COURSE CODE: 16B11BI112   MAX. MARKS: 35

COURSE NAME: Datawarehousing and Mining in Bioinformatics

COURSE CREDITS: 04   MAX. TIME: 2 Hr

Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means. Each question carry equal marks.

1. Why do we perform wavelet transformation on a dataset? (1)

2. How do we identify closed frequent itemsets and maximal frequent itemsets from a dataset? Illustrate with examples. (4)

3. Explain the meaning of significant and minor rules with respect to Li's committee approach of decision trees .How is this approach better than using single decision trees. (5)

4. Explain the role of eigen vectors and eigen values in Principal Components analysis. Describe the application of PCA to a biological example. (5)

5. How do we use Apriori algorithm to derive association rules? Substantiate your answer by giving examples. Also state how we can improve its efficiency. (5)

6. How can we use decision tree induction for feature selection on a dataset? Show with the help of a hypothetical decision tree. (3)

7. How do we identify whether an association rule is strong? Construct a hypothetical dataset and explain four different measures (with proper calculations) used to assess the strength of the rule. (4)

8. Discuss the three rules that we check as a part of the Data cleaning step in preprocessing. How are these useful?(3)

9. Explain why concept hierarchies are useful in data mining. (2)

10. Use a flowchart to summarize the following procedures for attribute subset selection: (3)
    (a) stepwise forward selection
    (b) stepwise backward elimination
    (c) a combination of forward selection and backward elimination